

A Superlinearly Convergent Polynomial
Primal-Dual Interior-Point Algorithm
for Linear Programming

Yin Zhang

and

R.A. Tapia

June, 1990

(Revised December, 1991)

TR90-40

Report Documentation Page				Form Approved OMB No. 0704-0188	
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE DEC 1991		2. REPORT TYPE		3. DATES COVERED 00-00-1991 to 00-00-1991	
4. TITLE AND SUBTITLE A Superlinearly Convergent Polynomial Primal-Dual Interior-Point Algorithm for Linear Programming				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Computational and Applied Mathematics Department ,Rice University,6100 Main Street MS 134,Houston,TX,77005-1892				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 31	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

A Superlinearly Convergent Polynomial Primal-Dual Interior-Point Algorithm for Linear Programming

Yin Zhang ^{*}and Richard A. Tapia [†]

June, 1990

(Revised December, 1991)

Key Words: Linear programming, Primal-dual interior-point algorithms, Polynomiality, Quadratic and superlinear convergence.

Abbreviated Title: A Superlinearly Convergent Polynomial Algorithm

^{*}Department of Mathematics and Statistics, University of Maryland, Baltimore County Campus, Baltimore, MD 21228, and visiting member of the Center for Research in Parallel Computation, Rice University. This author was supported in part by NSF Coop. Agr. No. CCR-8809615, NSF DMS-9102761 and DOE DE-FG05-91ER25100.

[†]Department of Mathematical Sciences and the Center for Research in Parallel Computation, Rice University, Houston, Texas 77251-1892. This author was supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.

Abstract

The choice of the centering parameter and the step-length parameter are the fundamental issues in primal-dual interior-point algorithms for linear programming. Various choices for these two parameters have been proposed that lead to polynomial algorithms. Recently, Zhang, Tapia and Dennis derived conditions for the choices of the two parameters that were sufficient for superlinear or quadratic convergence. However, prior to this work it had not been shown that these conditions for fast convergence are compatible with the choices that lead to polynomiality; none of the existing polynomial primal-dual interior-point algorithms satisfies these fast convergence requirements. This paper gives an affirmative answer to the question: can a primal-dual algorithm be both polynomial and superlinearly convergent for general problems? We construct and analyze a “large step” algorithm that possesses both polynomiality and, under the assumption of the convergence of the iteration sequence, Q -superlinear convergence. For nondegenerate problems, the convergence is actually Q -quadratic.

1 Introduction

We consider linear programs in the standard form:

$$\begin{aligned} & \text{minimize} && c^T x \\ & \text{subject to} && Ax = b, \\ & && x \geq 0, \end{aligned} \tag{1.1}$$

where $c, x \in \mathbf{R}^n$, $b \in \mathbf{R}^m$, $A \in \mathbf{R}^{m \times n}$ ($m < n$) and A is assumed to have full rank m .

The first-order optimality conditions for (1.1) can be written

$$\begin{pmatrix} Ax - b \\ A^T \lambda + y - c \\ XYe \end{pmatrix} = 0, \quad (x, y) \geq 0, \tag{1.2}$$

where λ and y are dual variables, $X = \text{diag}(x)$, $Y = \text{diag}(y)$ and e has all components equal to one. To facilitate our presentation, we will eliminate the dual variable λ from the above system (though such an elimination may not be advisable from a practical point of view). Let $B \in \mathbf{R}^{(n-m) \times n}$ be any matrix such that the columns of B^T form a basis for the null space of A . Pre-multiply the second equation by the nonsingular matrix $[A^T \ B^T]^T$. Notice that $BA^T = 0$, so

$$0 = \begin{bmatrix} A \\ B \end{bmatrix} (A^T \lambda + y - c) = \begin{pmatrix} AA^T \lambda + A(y - c) \\ By - Bc \end{pmatrix}.$$

Since AA^T is nonsingular, λ is uniquely determined once y is known. Removing the equation for λ , we arrive at the following $2n$ by $2n$ nonlinear system with non-negativity constraints on the variables:

$$F(x, y) = \begin{pmatrix} Ax - b \\ By - Bc \\ XYe \end{pmatrix} = 0, \quad (x, y) \geq 0. \quad (1.3)$$

By the feasibility set of problem (1.3) we mean:

$$\Omega = \{(x, y) : x, y \in \mathbf{R}^n, Ax = b, By = Bc, (x, y) \geq 0\}.$$

A feasible pair $(x, y) \in \Omega$ is said to be strictly feasible if it is positive. In this work we tacitly assume that strictly feasible points exist.

It is easy to see that for $(x, y) \in \Omega$, $\|F(x, y)\|_1 = x^T y$ which can be shown to be the duality gap for problem (1.1); we will use the duality gap as the merit function for our algorithm, i.e., the criterion that tells us when one feasible point should be preferred to another.

Mathematically speaking, the concepts of polynomiality and rate of convergence are incompatible. Polynomiality is meaningful only for algorithms that converge in a finite number of steps, while rate of convergence is defined only for algorithms that take an infinite number of steps to converge. When we say that an interior-point algorithm is

polynomial, we have in mind integral (or rational) data and finite termination. On the other hand, when we say the same algorithm is linearly convergent, for example, we do so in the traditional numerical analysis sense. With this understanding, we can discuss polynomiality and rate of convergence of an algorithm at the same time.

It is clearly desirable to develop algorithms that possess both polynomiality and fast asymptotic convergence, or in other words, both good global behavior and good local behavior. To our knowledge, the only prior work in this direction was Yamashita [10]. Based on the multiplicative penalty function of Iri and Imai [2], Yamashita constructed a polynomial primal algorithm and demonstrated its quadratic convergence under the following two assumptions: (i) the optimal objective value is known, and (ii) the iteration sequence converges to a nondegenerate optimal vertex. The first assumption is in general not realistic. The second assumption is very restrictive because most practical problems are degenerate.

The objective of this work is to construct a primal-dual interior-point algorithm for problem (1.1) that possesses both polynomiality and fast convergence under more realistic and less restrictive assumptions. We construct such an algorithm and show that it takes at most $O(nL)$ iterations to reduce the duality gap to 2^{-L} . Moreover, we demonstrate that this algorithm gives quadratic convergence for nondegenerate problems and gives Q -superlinear convergence for degenerate problems.

Subscripts will be used to distinguish values of quantities at a particular iteration and superscripts will indicate components of vectors. We also use the notation:

$$\min(v) = \min_{1 \leq i \leq n} v^i \quad \text{and} \quad \max(v) = \max_{1 \leq i \leq n} v^i$$

for a vector $v \in \mathbf{R}^n$. The symbol $\|\cdot\|$ denotes the ℓ_2 norm unless otherwise stated. We will use the standard big- O notation in this paper; in particular, for a sequence $\{v_k\} \subset \mathbf{R}^n$ and a positive sequence $\{\alpha_k\} \subset \mathbf{R}$, $v_k = O(\alpha_k)$ implies the existence of positive constants β and k_0 such that $\|v_k\| \leq \beta\alpha_k$ for all $k > k_0$.

The paper is organized as follows. In Section 2, we describe a general interior-point

algorithmic framework for problem (1.1) based on the nonlinear system (1.3) and give a brief survey of existing results for algorithms that fall into this framework. In Sections 3 and 4, we specify our procedures for determining the step length and for choosing the centering parameter. We state our algorithm in Section 5. Global linear convergence (and polynomiality) is established in Section 6. Quadratic convergence for nondegenerate problems is established in Section 7, and superlinear convergence for all problems is established in Section 8. Concluding remarks are given in Section 9.

2 General Algorithm

Algorithm 1 (General Algorithm)

Given a strictly feasible pair (x_0, y_0) . For $k = 0, 1, 2, \dots$, do

Step 1 *Compute the Newton step*

$$\begin{pmatrix} \Delta x_k^N \\ \Delta y_k^N \end{pmatrix} = -[F'(x_k, y_k)]^{-1} F(x_k, y_k)$$

and the centering step

$$\begin{pmatrix} \Delta x_k^C \\ \Delta y_k^C \end{pmatrix} = \frac{1}{n} x_k^T y_k [F'(x_k, y_k)]^{-1} \begin{pmatrix} 0 \\ e \end{pmatrix}.$$

Step 2 *Choose $\sigma_k \in (0, 1)$ and form the combined step*

$$\begin{pmatrix} \Delta x_k \\ \Delta y_k \end{pmatrix} = \begin{pmatrix} \Delta x_k^N \\ \Delta y_k^N \end{pmatrix} + \sigma_k \begin{pmatrix} \Delta x_k^C \\ \Delta y_k^C \end{pmatrix}.$$

Step 3 *Choose $\tau_k \in (0, 1)$ and set $\alpha_k = \tau_k \hat{\alpha}_k$, where*

$$\hat{\alpha}_k = \frac{-1}{\min(X_k^{-1} \Delta x_k, Y_k^{-1} \Delta y_k)}.$$

Step 4 *Compute the new iterate*

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \alpha_k \begin{pmatrix} \Delta x_k \\ \Delta y_k \end{pmatrix}.$$

We will now briefly comment on this general algorithmic framework. From a direct calculation, we have

$$F'(x, y) = \begin{bmatrix} A & 0 \\ 0 & B \\ Y & X \end{bmatrix}. \quad (2.1)$$

Since we have assumed that A has full rank, it is a straightforward matter to verify that $F'(x, y)$ is nonsingular for any positive pair (x, y) . In addition, relation (2.5) below guarantees that $\hat{\alpha}_k > 0$. Hence the iterates produced by Algorithm 1 are well-defined. Notice that the restriction $\alpha_k < \hat{\alpha}_k$ guarantees that the iterates remain strictly feasible. Moreover, we have the following useful relationships:

$$Y_k \Delta x_k^N + X_k \Delta y_k^N = -X_k Y_k e, \quad (2.2)$$

$$Y_k \Delta x_k^C + X_k \Delta y_k^C = \frac{1}{n} x_k^T y_k e, \quad (2.3)$$

$$Y_k \Delta x_k + X_k \Delta y_k = -X_k Y_k e + \sigma_k \frac{1}{n} x_k^T y_k e, \quad (2.4)$$

$$\Delta x_k^T \Delta y_k = 0, \quad (2.5)$$

$$x_{k+1}^T y_{k+1} = x_k^T y_k (1 - (1 - \sigma_k) \alpha_k). \quad (2.6)$$

We have stated Algorithm 1 in this form for notational convenience. It is not difficult to verify that identical iterates $\{(x_k, y_k)\}$ can be generated using (1.2) instead of (1.3). For this case, there is no need for introducing the matrix B (see [11] for example).

From (2.6) we see that Algorithm 1 is a descent algorithm for the duality gap $\|F(x, y)\|_1 = x^T y$. Moreover, the duality gap is reduced at iteration k by a factor $1 - \alpha_k(1 - \sigma_k) < 1$; thus, linear convergence will be obtained if $\{\alpha_k\}$ is bounded away from zero and $\{\sigma_k\}$ is bounded away from one. In addition, Q -superlinear convergence will be obtained if $\alpha_k(1 - \sigma_k) \rightarrow 1$. Observe that we have direct control over the choice of

σ_k . However, we do not have the freedom of choosing α_k uniformly bounded away from zero, since we must enforce the requirement $\alpha_k < \hat{\alpha}_k$ and $\hat{\alpha}_k$ is not directly under our control.

A number of existing primal-dual algorithms fit into the above general algorithmic framework with different choices for the parameters σ_k and α_k . For example, in the primal-dual algorithm of Kojima, Mizuno and Yoshise [3], σ_k is a constant and α_k is a particular function of σ_k . They showed that their algorithm requires at most $O(nL)$ iterations to reduce the duality gap by a factor of 2^{-L} . Other examples include the Todd and Ye [9] primal-dual potential reduction algorithm and the Monteiro and Adler [7] path-following primal-dual algorithm. Todd and Ye's algorithm uses the choice

$$\sigma_k = \frac{\sqrt{n}}{\sqrt{n} + \nu}$$

where ν is a constant. In Monteiro and Adler's algorithm,

$$\sigma_k = 1 - \frac{\delta}{\sqrt{n}}$$

where δ is a constant (Monteiro and Adler actually used $\delta = 0.35$ in their analysis). In both algorithms, a rather short step-length α_k is required. Furthermore, both of these algorithms require at most $O(\sqrt{n}L)$ iterations to reduce the duality gap to 2^{-L} . This is the best complexity bound obtained for linear programming so far. Observe that all three algorithms use constant σ_k . In each of the three cases if σ denotes the constant value of σ_k , then Q -superlinear convergence is possible (see (2.6)) only if

$$\alpha_k \rightarrow \frac{1}{1 - \sigma},$$

which seems extremely unlikely.

In analyzing the convergence of Algorithm 1, a central quantity is

$$\eta_k = \frac{x_k^T y_k / n}{\min(X_k Y_k e)}. \quad (2.7)$$

Since $\frac{1}{n}x_k^T y_k$ is the average value of the components of $X_k Y_k e$, it is clear that $\eta_k \geq 1$. In all the above mentioned polynomial algorithms, it is essential that the sequence $\{\eta_k\}$ be bounded.

Recently, Zhang, Tapia and Dennis [11] showed that under appropriate assumptions, Algorithm 1 has fast convergence. The following two theorems summarize their main results. By a nondegenerate vertex of (1.1), we mean a feasible point of (1.1) that has exactly m positive components and the corresponding m columns of A are linearly independent.

Theorem 2.1 (Zhang, Tapia and Dennis)

Let (x_, y_*) be a solution of problem (1.3) and $\{(x_k, y_k)\}$ be generated by Algorithm 1. Assume*

- (i) *strict complementarity holds at (x_*, y_*) ,*
- (ii) *x_* is a nondegenerate vertex of (1.1),*
- (iii) *$\sigma_k = O(x_k^T y_k)$ and $\tau_k = 1 - O(x_k^T y_k)$.*

If $\{(x_k, y_k)\}$ converges to (x_, y_*) , then the convergence is Q -quadratic.*

Theorem 2.2 (Zhang, Tapia and Dennis)

Let (x_, y_*) be a solution of problem (1.3) and $\{(x_k, y_k)\}$ be generated by Algorithm 1. Assume*

- (i) *strict complementarity holds at (x_*, y_*) ,*
- (ii) *the sequence $\{\eta_k\}$ is bounded,*
- (iii) *$\sigma_k \rightarrow 0$ and $\tau_k \rightarrow 1$.*

If $\{(x_k, y_k)\}$ converges to (x_, y_*) , then the duality gap sequence $\{x_k^T y_k\}$ converges to zero Q -superlinearly.*

With some additional work, one can actually demonstrate that the sequence $\{X_k Y_k e\}$ component-wise converges to zero Q -superlinearly.

Several assumptions have been made in the above theorems. Our numerical experiments have led us to believe that the strict complementarity assumption is not restrictive. On the other hand, the nondegeneracy assumption is quite restrictive since degeneracy exists in most real-world problems. For degenerate solutions, the best convergence that has been established is Q -superlinear, as stated in Theorem 2.2.

Although many of the existing polynomial primal-dual interior-point algorithms satisfy assumption (ii) of Theorem 2.2, none of them satisfy assumption (iii), i.e., $\sigma_k \rightarrow 0$ and $\tau_k \rightarrow 1$. In fact, in several polynomial algorithms, for example Todd and Ye's and Monteiro and Adler's, the values of σ_k are close to one. From Zhang, Tapia and Dennis [11] it follows that these algorithms will most likely have slow Q -linear convergence. Hence while their global behavior may be excellent, their local behavior can be improved.

Recently, in a number of performance-oriented primal-dual algorithms, for example the ones implemented by Choi *et. al.* [1], McShane *et. al.* [6] and Lustig *et. al.* [5], very small values of σ_k were used and also long steps were taken. Impressive numerical results were obtained from these implementations though polynomial complexity bounds are not known. Hence while their local behavior may be good, their global behavior is in question from a theoretical standpoint.

In this work, we develop a primal-dual interior-point polynomial algorithm that gives quadratic convergence for nondegenerate solutions and superlinear convergence for degenerate solutions. Hence, from a mathematical point of view, both the global and the local behavior will be good. This new algorithm is still of a theoretical nature. However, the fact that polynomiality and quadratic or superlinear convergence can be achieved simultaneously by one algorithm provides motivation for practical implementations of the conditions $\sigma_k = O(x_k^T y_k)$ and $\tau_k = 1 - O(x_k^T y_k)$ for fast convergence.

3 Determining the Step-Length

In the previous section we mentioned that both polynomiality and superlinear convergence essentially require that the sequence $\{\eta_k\}$ be bounded. The most straightforward way of accomplishing this objective is to explicitly enforce a uniform bound on the quantity

$$\eta_{k+1} = \frac{x_{k+1}^T y_{k+1} / n}{\min(X_{k+1} Y_{k+1} e)}$$

during the process of choosing the step-length α_k ; i.e., ask that

$$\frac{1}{\eta_{k+1}} = \frac{\min(X_{k+1} Y_{k+1} e)}{x_{k+1}^T y_{k+1} / n} \geq \gamma \quad (3.1)$$

for some $\gamma > 0$.

Following the notation used in [3], let

$$\begin{aligned} x_k(\alpha) &= x_k + \alpha \Delta x_k, & y_k(\alpha) &= y_k + \alpha \Delta y_k, \\ f_k(\alpha) &= X_k(\alpha) Y_k(\alpha) e, & f_k^{\text{ave}}(\alpha) &= \frac{1}{n} x_k(\alpha)^T y_k(\alpha), \\ f_k^{\min}(\alpha) &= \min(f_k(\alpha)), & f_k^{\max}(\alpha) &= \max(f_k(\alpha)). \end{aligned} \quad (3.2)$$

Note that the above quantities actually also depend on the centering parameter σ because both Δx_k and Δy_k are functions of σ (see Step 3 of Algorithm 1). However, since we will always choose σ before we determine α , it will suffice to consider these quantities only as functions of α for a fixed value of σ .

Whenever $\alpha = 0$, we will drop the argument from the above functions. For example, $x_k \equiv x_k(0)$, $f_k^{\text{ave}} \equiv f_k^{\text{ave}}(0)$ and so on. From the formula for the iterates (Step 4 of Algorithm 1), we also have $x_{k+1} = x_k(\alpha_k)$, $f_{k+1}^{\text{ave}} = f_k^{\text{ave}}(\alpha_k)$ and so on.

Using the above notation, we choose the form of condition (3.1) as requiring α_k to satisfy

$$\frac{f_k^{\min}(\alpha)}{f_k^{\text{ave}}(\alpha)} \geq \gamma_k, \quad \alpha > 0, \quad (3.3)$$

where

$$\gamma_k \in [\gamma, f_k^{\min} / f_k^{\text{ave}}], \quad 0 < \gamma \leq f_0^{\min} / f_0^{\text{ave}} \leq 1 \quad \text{and} \quad \gamma, \gamma_k < 1. \quad (3.4)$$

In the case $1/\eta_0 > \gamma$, we allow $1/\eta_k$ to decrease monotonically as long as $1/\eta_k > \gamma$.

In the following development, we will use some of the techniques developed by Kojima *et al* in [3].

Using (3.2), (2.4), (2.5), (2.6) and letting

$$s_k = \text{diag}(\Delta x_k) \Delta y_k,$$

we have

$$f_k^i(\alpha) = f_k^i - (f_k^i - \sigma_k f_k^{\text{ave}}) \alpha + s_k^i \alpha^2 \quad (3.5)$$

and

$$f_k^{\text{ave}}(\alpha) = f_k^{\text{ave}}[1 - (1 - \sigma_k) \alpha]. \quad (3.6)$$

Hence, $f_k^i(\alpha)$ is a quadratic (so $f_k^{\min}(\alpha)$ and $f_k^{\max}(\alpha)$ are piecewise quadratic) and $f_k^{\text{ave}}(\alpha)$ is linear. Clearly, if $f_k^{\text{ave}}(\hat{\alpha}_k) = 0$, then $(x_k(\hat{\alpha}_k), y_k(\hat{\alpha}_k))$ will solve problem (1.3). In the sequel, we always assume $f_k^{\text{ave}}(\hat{\alpha}_k) > 0$.

For notational convenience, let us introduce the piecewise quadratic function

$$h(\alpha) \stackrel{\text{def}}{=} f_k^{\min}(\alpha) - \gamma_k f_k^{\text{ave}}(\alpha). \quad (3.7)$$

It follows that condition (3.3) is equivalent to

$$h(\alpha) \geq 0, \quad \alpha > 0. \quad (3.8)$$

In determining α_k we will use the following quantity:

$$\alpha_k^\gamma \stackrel{\text{def}}{=} \min\{\alpha > 0 : h(\alpha) = 0\}. \quad (3.9)$$

Recall that $\hat{\alpha}_k$ is defined in Step 3 of the general algorithm (see Section 2).

Lemma 3.1 *The quantity α_k^γ is well defined and $\alpha_k^\gamma \in (0, \hat{\alpha}_k)$. Moreover, condition (3.3) is satisfied for all $\alpha \in (0, \alpha_k^\gamma]$.*

Proof: Let us examine the function $h(\alpha)$. It follows from the definitions of γ_k and $\hat{\alpha}_k$ that

$$h(0) = f_k^{\min} - \gamma_k f_k^{\text{ave}} \geq 0$$

and

$$h(\hat{\alpha}_k) = f_k^{\min}(\hat{\alpha}_k) - \gamma_k f_k^{\text{ave}}(\hat{\alpha}_k) = -\gamma_k f_k^{\text{ave}}(\hat{\alpha}_k) < 0.$$

Hence it follows from the continuity of $h(\alpha)$ that $h(\alpha)$ has a root in $[0, \hat{\alpha}_k)$. When $h(0) > 0$, $h(\alpha)$ obviously has a root in $(0, \hat{\alpha}_k)$. When $h(0) = 0$, it can be verified that the right-derivative of $h(\alpha)$ at $\alpha = 0$ is

$$\begin{aligned} h'(0^+) &= -(f_k^{\min} - \sigma_k f_k^{\text{ave}}) + \gamma_k(1 - \sigma_k)f_k^{\text{ave}} \\ &= [(1 - \gamma_k)\sigma_k + (\gamma_k - f_k^{\min}/f_k^{\text{ave}})]f_k^{\text{ave}} \\ &= (1 - \gamma_k)\sigma_k > 0. \end{aligned}$$

Therefore, $h(\alpha) > 0$ for sufficiently small but positive α . Consequently, $\alpha_k^\gamma > 0$.

Since $h(\hat{\alpha}_k) < 0$, we have $\alpha_k^\gamma < \hat{\alpha}_k$. It is evident that $h(\alpha) \geq 0$ for $\alpha \in (0, \alpha_k^\gamma]$, i.e., condition 3.3 is satisfied. This completes the proof. \square

An equivalent expression for α_k^γ is

$$\alpha_k^\gamma = \min\{\alpha > 0 : f_k^i(\alpha) - \gamma_k f_k^{\text{ave}}(\alpha) = 0, i = 1, 2, \dots, n\}. \quad (3.10)$$

The computation of α_k^γ involves calculating the roots of at most n quadratics and therefore requires $O(n)$ operations.

In addition to a lower bound for $\{f_k^i(\alpha_k)/f_k^{\text{ave}}(\alpha_k)\}$ (i.e., condition (3.3)), we also impose an upper bound on these quantities; namely, we require α_k to satisfy

$$\frac{f_k^{\max}(\alpha)}{f_k^{\text{ave}}(\alpha)} \leq \Gamma_k, \quad \alpha > 0, \quad (3.11)$$

where

$$\Gamma_k \in [f_k^{\max}/f_k^{\text{ave}}, \Gamma], \quad 1 \leq f_0^{\max}/f_0^{\text{ave}} \leq \Gamma \leq n \quad \text{and} \quad \Gamma, \Gamma_k > 1. \quad (3.12)$$

Since $f_k^i(\alpha)/f_k^{\text{ave}}(\alpha) < n$ for all i , condition (3.11) will be redundant if $\Gamma_k = n$. Our reason for introducing condition (3.11) is to improve our complexity bound. We do not feel that enforcing this condition will have much practical significance.

Following the treatment of condition (3.3), we introduce the piecewise quadratic function

$$H(\alpha) \stackrel{\text{def}}{=} f_k^{\max}(\alpha) - \Gamma_k f_k^{\text{ave}}(\alpha). \quad (3.13)$$

It follows that condition (3.11) is equivalent to

$$H(\alpha) \leq 0, \quad \alpha > 0. \quad (3.14)$$

We will also use the following quantity in determining α_k :

$$\alpha_k^\Gamma \stackrel{\text{def}}{=} \min\{\alpha > 0 : H(\alpha) = 0\}. \quad (3.15)$$

Analogous to Lemma 3.1 for condition (3.3), we have the following lemma for condition (3.11).

Lemma 3.2 *The quantity α_k^Γ is well-defined and $\alpha_k^\Gamma \in (0, \hat{\alpha}_k)$. Moreover, condition (3.11) is satisfied by all $\alpha \in (0, \alpha_k^\Gamma]$.*

Proof: The proof is similar to that for Lemma 3.1, so we omit it. \square

Analogous to the expression (3.10) for condition (3.3), we have for condition (3.11)

$$\alpha_k^\Gamma = \min\{\alpha > 0 : f_k^i(\alpha) - \Gamma_k f_k^{\text{ave}}(\alpha) = 0, \quad i = 1, 2, \dots, n\}. \quad (3.16)$$

For the sake of simplicity, we will enforce the conditions

$$\gamma_k \leq 1/2 \quad \text{and} \quad \Gamma_k \geq 2. \quad (3.17)$$

The specific values in (3.17) do not constitute a loss of generality because they will only affect expressions for some constants in our analysis. These values of γ_k and Γ_k will result in much simplified expressions for those constants.

From (2.2), we see that for fixed σ_k a larger step length α_k will produce a larger reduction in the duality gap. So it is always desirable to take the largest step-length possible as long as other requirements are satisfied. Our procedure for determining the step-length α_k is summarized as follows.

Procedure 1 (Step-length Criterion)

Given positive constants γ and Γ such that

$$0 < \gamma \leq \min(1/2, f_0^{\min}/f_0^{\text{ave}}), \quad \max(2, f_0^{\max}/f_0^{\text{ave}}) \leq \Gamma < n. \quad (3.18)$$

Step 1 Choose $\gamma_k \in [\gamma, \min(1/2, f_k^{\min}/f_k^{\text{ave}})]$ and $\Gamma_k \in [\max(2, f_k^{\max}/f_k^{\text{ave}}), \Gamma]$.

Step 2 Compute $\alpha_k^\gamma = \min\{\alpha > 0 : f_k^i(\alpha) - \gamma_k f_k^{\text{ave}}(\alpha) = 0, i = 1, 2, \dots, n\}$ (i.e., (3.9)).

Step 3 Compute $\alpha_k^\Gamma = \min\{\alpha > 0 : f_k^i(\alpha) - \Gamma_k f_k^{\text{ave}}(\alpha) = 0, i = 1, 2, \dots, n\}$ (i.e., (3.15)).

Step 4 Let $\alpha_k = \min(\alpha_k^\gamma, \alpha_k^\Gamma)$.

We notice that the above procedure of choosing the step-length bears a certain similarity to a procedure recently proposed by Mizuno, Todd and Ye [8].

Now we prove two technical lemmas that will be needed in the later development.

Lemma 3.3 For $\alpha \in [0, 1]$,

$$\begin{aligned} f_k^{\min}(\alpha) &\geq f_k^{\min} - (f_k^{\min} - \sigma_k f_k^{\text{ave}})\alpha + \min(s_k)\alpha^2, \\ f_k^{\max}(\alpha) &\leq f_k^{\max} - (f_k^{\max} - \sigma_k f_k^{\text{ave}})\alpha + \max(s_k)\alpha^2. \end{aligned}$$

Proof: We first look at the linear part of $f_k^i(\alpha)$. Since for all i ,

$$f_k^i - (f_k^i - \sigma_k f_k^{\text{ave}})\alpha = \begin{cases} f_k^i, & \alpha = 0, \\ \sigma_k f_k^{\text{ave}}, & \alpha = 1, \end{cases}$$

it is evident that for $\alpha \in [0, 1]$

$$f_k^{\min} - (f_k^{\min} - \sigma_k f_k^{\text{ave}})\alpha \leq f_k^i - (f_k^i - \sigma_k f_k^{\text{ave}})\alpha \leq f_k^{\max} - (f_k^{\max} - \sigma_k f_k^{\text{ave}})\alpha.$$

For the quadratic terms, we clearly have

$$\min(s_k)\alpha^2 \leq s_k^i \alpha^2 \leq \max(s_k)\alpha^2.$$

By adding the quadratic terms to their corresponding linear parts, we finish the proof.

□

It is worth noting that $e^T s_k = 0$ by (2.5). Hence, $\min(s_k) \leq 0$ and $\max(s_k) \geq 0$. In the sequel, we will adopt the convention that $\frac{1}{0} = +\infty$.

Lemma 3.4 *Let α_k be given by Procedure 1. Then*

$$\alpha_k \geq \min \left(1, \frac{(1 - \gamma_k) \sigma_k f_k^{\text{ave}}}{-\min(s_k)}, \frac{(\Gamma_k - 1) \sigma_k f_k^{\text{ave}}}{\max(s_k)} \right). \quad (3.19)$$

Moreover,

$$\alpha_k \geq \min \left(1, \frac{\sigma_k f_k^{\text{ave}}}{2 \|s_k\|_\infty} \right). \quad (3.20)$$

Proof: From (3.9), α_k^γ is a positive root of $f_k^i(\alpha) - \gamma_k f_k^{\text{ave}}(\alpha)$ for some index i . Noticing that for $\alpha \in [0, 1]$ $f_k^{\text{ave}}(\alpha)$ is positive, and using Lemma 3.3, for $\alpha \in [0, 1]$, $\gamma_k \geq 0$ and for all index i , we have

$$\begin{aligned} f_k^i(\alpha) - \gamma_k f_k^{\text{ave}}(\alpha) &\geq f_k^{\min} - (f_k^{\min} - \sigma_k f_k^{\text{ave}}) \alpha + \min(s_k) \alpha^2 - \gamma_k f_k^{\text{ave}}(\alpha) \\ &= (f_k^{\min} - \gamma_k f_k^{\text{ave}})(1 - \alpha) + (1 - \gamma_k) \sigma_k f_k^{\text{ave}} \alpha + \min(s_k) \alpha^2 \\ &\geq (1 - \gamma_k) \sigma_k f_k^{\text{ave}} \alpha + \min(s_k) \alpha^2. \end{aligned} \quad (3.21)$$

If $\min(s_k) = 0$, then $h(\alpha) > 0$ for $\alpha \in (0, 1]$. Therefore, we will have $\alpha_k^\gamma > 1$. Now assume $\min(s_k) < 0$. Then the quadratic on the right-hand side of the last inequality in (3.21) has a unique positive root

$$\bar{\alpha}_k = \frac{(1 - \gamma_k) \sigma_k f_k^{\text{ave}}}{-\min(s_k)}.$$

Hence, if $\alpha_k^\gamma \leq 1$, from (3.21) we must have $\alpha_k^\gamma \geq \bar{\alpha}_k$. This proves that

$$\alpha_k^\gamma \geq \min \left(1, \frac{(1 - \gamma_k) \sigma_k f_k^{\text{ave}}}{-\min(s_k)} \right). \quad (3.22)$$

Similarly, we can prove that

$$\alpha_k^\Gamma \geq \min \left(1, \frac{(\Gamma_k - 1) \sigma_k f_k^{\text{ave}}}{\max(s_k)} \right). \quad (3.23)$$

Combining (3.22) and (3.23), we obtain (3.19).

Finally, (3.20) follows from the facts that $\|s_k\|_\infty = \max\{-\min(s_k), \max(s_k)\}$ and

$$\frac{1}{2} \leq 1 - \gamma_k < 1 \leq \Gamma_k - 1.$$

This completes the proof. \square

4 Choosing the Centering Parameter

We will use the following notation:

$$\begin{aligned} p_k &= X_k^{-1} \Delta x_k, \quad q_k = Y_k^{-1} \Delta y_k, \\ p_k^N &= X_k^{-1} \Delta x_k^N, \quad q_k^N = Y_k^{-1} \Delta y_k^N, \\ p_k^C &= X_k^{-1} \Delta x_k^C, \quad q_k^C = Y_k^{-1} \Delta y_k^C, \end{aligned} \tag{4.1}$$

and

$$\omega_k = \max_{1 \leq i \leq n} (|(p_k^N)^i (q_k^N)^i|, |(p_k^N)^i (q_k^C)^i|, |(p_k^C)^i (q_k^N)^i|, |(p_k^C)^i (q_k^C)^i|). \tag{4.2}$$

Lemma 4.1 *If $f_k^{\min}/f_k^{\text{ave}} \geq \gamma$, then*

$$\omega_k \leq n/\gamma^2.$$

Proof: Multiply both sides of (2.3) by $(X_k Y_k)^{-\frac{1}{2}}$ and consider the square of the ℓ_2 -norm of both sides. Using (2.5) and (4.1), we obtain

$$\|(X_k Y_k)^{\frac{1}{2}} p_k^C\|_2^2 + \|(X_k Y_k)^{\frac{1}{2}} q_k^C\|_2^2 = (\frac{1}{n} x_k^T y_k)^2 e^T (X_k Y_k)^{-1} e;$$

or equivalently after dividing both sides by $\frac{1}{n} x_k^T y_k$,

$$\|T_k^{-\frac{1}{2}} p_k^C\|_2^2 + \|T_k^{-\frac{1}{2}} q_k^C\|_2^2 = e^T T_k e, \tag{4.3}$$

where $T_k = \frac{1}{n} x_k^T y_k (X_k Y_k)^{-1}$ is a diagonal matrix. Our assumption implies that the maximum diagonal element of $\{T_k\}$ is bounded above by $1/\gamma$ and the minimum diagonal element of $\{T_k^{-1}\}$ is bounded below by γ . Therefore, from (4.3) we have

$$|(p_k^C)^i| \leq \sqrt{n}/\gamma \quad \text{and} \quad |(q_k^C)^i| \leq \sqrt{n}/\gamma.$$

Using the same technique, we can prove that

$$|(p_k^N)^i| \leq \sqrt{n/\gamma} \leq \sqrt{n}/\gamma \quad \text{and} \quad |(q_k^N)^i| \leq \sqrt{n/\gamma} \leq \sqrt{n}/\gamma.$$

From the definition of ω_k and the above estimates, Lemma 4.1 follows immediately. \square

We now state our procedure for choosing the centering parameter σ_k .

Procedure 2 (Centering Parameter Criterion)

Given

$$\sigma \in (0, 1), \quad \rho^l = \frac{\gamma^2 \sigma}{2n}, \quad \rho^u \geq \frac{\gamma^2 \sigma}{n}. \quad (4.4)$$

Step 1 Compute ω_k from (4.2).

Step 2 Compute $\rho_k^u = \min(\rho^u, \sigma/\omega_k)$.

Step 3 Choose $\rho_k \in [(\rho^l + \rho_k^u)/2, \rho_k^u]$.

Step 4 Let $\sigma_k = \rho_k \omega_k$.

Since $\sigma_k = \rho_k \omega_k$ and $\rho_k \in [\rho^l, \rho_k^u]$, we have $\sigma_k \in [\rho^l \omega_k, \rho_k^u \omega_k]$. In addition, we require that σ_k be greater than the midpoint of the interval. This requirement is needed in our proof of superlinear convergence. It is evident that σ_k is bounded away from one because $\sigma_k \leq \sigma < 1$. The reasons why the centering parameter is so chosen will hopefully become clear as our discussion proceeds.

5 Algorithm Description

Now we formally state our primal-dual interior-point algorithm.

Algorithm 2 Suppose given a strictly feasible pair (x_0, y_0) . Choose positive constants γ and Γ such that (see (3.18))

$$0 < \gamma \leq \min(1/2, f_0^{\min}/f_0^{\text{ave}}), \quad \max(2, f_0^{\max}/f_0^{\text{ave}}) \leq \Gamma < n,$$

and choose $\sigma \in (0, 1)$. Set $\rho^l = \gamma^2 \sigma / 2n$ and $\rho^u \geq \gamma^2 \sigma / n$ (see (4.4)). For $k = 0, 1, 2, \dots$, do

Step 1 Compute the Newton step and the centering step from Algorithm 1.

Step 2 Choose σ_k by Procedure 2 and form $(\Delta x_k, \Delta y_k)$ from Algorithm 1.

Step 3 Choose α_k by Procedure 1.

Step 4 Form (x_{k+1}, y_{k+1}) from Algorithm 1.

The procedure for determining the step-length α_k can be implemented in an effective manner. Its cost is somewhat higher than the ratio test that is used in most of the practical implementations. On the other hand, our procedure for choosing the centering parameter σ_k requires extra work when compared to the more standard method. The standard practice is to choose the centering parameter prior to computing the steps, then one only needs to solve once for the combined step (Newton step plus the centering parameter times the centering step). Since Algorithm 2 requires the information obtained from the Newton step and the centering step to choose the centering parameter, it requires one to solve for the two steps separately and then combine them.

6 Global Linear Convergence

Theorem 6.1 (Global Linear Convergence)

Let $\{(x_k, y_k)\}$ be generated by Algorithm 2. Then

$$x_{k+1}^T y_{k+1} \leq (1 - \delta/n) x_k^T y_k,$$

for some δ satisfying

$$\delta \geq \frac{\sigma(1 - \sigma)\gamma^2}{16\Gamma}.$$

Proof: We need to estimate $\|s_k\|_\infty$ in (3.20). Let the index j be such that $\|s_k\|_\infty = |s_k^j|$. Observe that

$$\begin{aligned}
\|s_k\|_\infty &= |\Delta x_k^j \Delta y_k^j| = |(x_k^j p_k^j)(y_k^j q_k^j)| = |(x_k^j y_k^j)(p_k^j q_k^j)| \\
&\leq \max(X_k Y_k e) \|\text{diag}(p_k) q_k\|_\infty \\
&= f_k^{\max} \|\text{diag}(p_k^N + \sigma_k p_k^C)(q_k^N + \sigma_k p_k^C)\|_\infty \\
&\leq f_k^{\max} \omega_k (1 + \sigma_k)^2 \\
&\leq 4 f_k^{\max} \omega_k.
\end{aligned}$$

Hence, it follows from (3.11), (3.20) and Procedure 2 that

$$\alpha_k \geq \min \left(1, \frac{\rho_k f_k^{\text{ave}}}{8 f_k^{\max}} \right) \geq \min \left(1, \frac{\rho_k}{8 \Gamma} \right) \geq \frac{\rho^l}{8 \Gamma}. \quad (6.1)$$

Substituting ρ^l (see (4.4)) into the above expression, we obtain

$$\alpha_k \geq \frac{\sigma \gamma^2}{16 \Gamma n}.$$

The proof is completed by substituting the above inequality into (2.6) and noticing that $\sigma_k \leq \sigma$. \square

The following corollary follows immediately from Theorem 6.1. By a standard argument, it leads to polynomiality assuming integral data.

Corollary 6.1 *Assume that a strictly feasible pair (x_0, y_0) , constants γ and Γ , both independent of n , are chosen such that (3.18) is satisfied and $x_0^T y_0 \leq 2^{\nu L}$, where $L > 0$ and ν is a positive constant independent of n . Then in at most $O(nL)$ iterations, Algorithm 2 will produce (x_k, y_k) such that $x_k^T y_k \leq 2^{-L}$.*

Proof: From Theorem 6.1,

$$x_k^T y_k \leq (1 - \delta/n)^k x_0^T y_0 \leq (1 - \delta/n)^k 2^{\nu L}.$$

Let $(1 - \delta/n)^k 2^{\nu L} = 2^{-L}$ and take the natural logarithm of both sides. We have $k = -(\ln 2)(1 + \nu)L / \ln(1 - \delta/n)$. Observe that for $x \in (0, 1)$

$$\ln(1 - x) = - \sum_{k=1}^{\infty} \frac{x^k}{k} < -x.$$

Therefore,

$$k \leq (\ln 2)(1 + \nu)L/(\delta/n) = O(nL).$$

This completes the proof. \square

7 Quadratic Convergence

In this section, we will apply Theorem 2.1 to establish that under strict complementarity and nondegeneracy assumptions our algorithm converges Q -quadratically. It can be shown that the nondegeneracy and strict complementarity assumptions at optimality imply the uniqueness of both primal and dual solutions. We have already established convergence of the duality gap sequence to zero in the preceding section. With the uniqueness, it can be shown that the convergence of the duality gap implies that of the iterates to the unique solution $(x_*, y_*) \geq 0$. What we must verify is assumption (iii) of Theorem 2.1; namely,

$$\sigma_k = O(x_k^T y_k) \quad \text{and} \quad \tau_k = 1 - O(x_k^T y_k).$$

Since $\tau_k = \alpha_k / \hat{\alpha}_k$, for the latter it suffices to show that

$$\hat{\alpha}_k \rightarrow 1 \quad \text{and} \quad \hat{\alpha}_k - \alpha_k = O(x_k^T y_k). \tag{7.1}$$

The following lemma will be useful. It is a slightly modified version of Lemma 3.2 in [11]. We refer interested readers to the original paper for its proof.

Lemma 7.1 (Zhang, Tapia and Dennis)

Let (x_, y_*) be a solution of problem (1.3) and $\{(x_k, y_k)\}$ be generated by Algorithm 2. Let p_k^N, p_k^C, q_k^N and q_k^C be defined by (4.1). Assume*

- (i) *strict complementarity holds at (x_*, y_*) ,*
- (ii) *x_* is a nondegenerate vertex of (1.1),*

Then

$$p_k^N = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ -1 \\ \vdots \\ -1 \end{pmatrix} + O(x_k^T y_k), \quad p_k^C = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \frac{x_k^T y_k / n}{[X_k Y_k e]^{m+1}} \\ \vdots \\ \frac{x_k^T y_k / n}{[X_k Y_k e]^n} \end{pmatrix} + O(x_k^T y_k),$$

and

$$q_k^N = \begin{pmatrix} -1 \\ \vdots \\ -1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + O(x_k^T y_k), \quad q_k^C = \begin{pmatrix} \frac{x_k^T y_k / n}{[X_k Y_k e]^1} \\ \vdots \\ \frac{x_k^T y_k / n}{[X_k Y_k e]^m} \\ 0 \\ \vdots \\ 0 \end{pmatrix} + O(x_k^T y_k),$$

where the number of zeros is m in p_k^N and p_k^C , and $n - m$ in q_k^N and q_k^C .

Now we are ready to state and prove our quadratic convergence theorem.

Theorem 7.1 (Quadratic Convergence)

Let (x_*, y_*) be a solution of problem (1.3) and $\{(x_k, y_k)\}$ be generated by Algorithm 2. Assume

- (i) strict complementarity holds at (x_*, y_*) ,
- (ii) x_* is a nondegenerate vertex of (1.1),
- (iii) ρ^u is sufficiently large, e.g., $\rho^u \geq 16\Gamma$.

Then $\{(x_k, y_k)\}$ converges to (x_*, y_*) Q -quadratically.

Proof: We first prove $\sigma_k = O(x_k^T y_k)$. Observe from Lemma 7.1 that for each index i either the “ p ” terms $((p_k^N)^i$ and $(p_k^C)^i)$ or the “ q ” terms $((q_k^N)^i$ and $(q_k^C)^i)$ are $O(x_k^T y_k)$

while the other terms are bounded. Thus, the quantity ω_k (see its definition (4.2)) is $O(x_k^T y_k)$. So is σ_k because $\sigma_k \leq \rho^u \omega_k$.

Since $\omega_k \rightarrow 0$, from the choice of ρ_k^u in Step 2 of Procedure 2 we have for k sufficiently large

$$\rho_k^u = \rho^u \quad \text{and} \quad \rho_k \geq \frac{1}{2}(\rho^l + \rho^u). \quad (7.2)$$

We observe that if ρ^u is sufficiently large, e.g., $\rho^u \geq 16\Gamma$, (i.e., σ_k is not forced to approach zero too quickly), then the step-length α_k will eventually be equal to or greater than one, as can be seen from (6.1).

Since $\sigma_k = O(x_k^T y_k)$ and $(x_k^T y_k/n)/\min(X_k Y_k e)$ is bounded, the elements of p_k and q_k are either $O(x_k^T y_k)$ or $-1 + O(x_k^T y_k)$. Therefore,

$$\min(X_k^{-1} \Delta x_k, Y_k^{-1} \Delta y_k) = \min(p_k, q_k) = -1 + O(x_k^T y_k). \quad (7.3)$$

By examining the definition of $\hat{\alpha}_k$ in Step (3) of Algorithm 1, we see $\hat{\alpha}_k = 1 + O(x_k^T y_k)$. Consequently, for k sufficiently large we have

$$1 \leq \alpha_k < \hat{\alpha}_k = 1 + O(x_k^T y_k).$$

This implies (7.1) and completes the proof. □

8 Superlinear Convergence

In this section, we will apply Theorem 2.2 to establish Q -superlinear convergence of Algorithm 2 for general problems. We must show that assumption (iii) of Theorem 2.2 holds; i.e.,

$$\sigma_k \rightarrow 0 \quad \text{and} \quad \tau_k \rightarrow 1.$$

For the latter, it will suffice to show $\hat{\alpha}_k \rightarrow 1$ and $\hat{\alpha}_k - \alpha_k \rightarrow 0$. Without the nondegeneracy assumption, we can no longer use Lemma 7.1. For technical reasons, we must restrict further the choice of ρ_k .

Denote the length of the interval $[\rho^l, \rho_k^u]$ by π_k . It follows from (4.4), Step 2 of Procedure 2, and Lemma 4.1 that

$$\rho_k^u \geq \frac{\gamma^2 \sigma}{n}. \quad (8.1)$$

Thus,

$$\pi_k \stackrel{\text{def}}{=} \rho_k^u - \rho^l \geq \frac{\gamma^2 \sigma}{2n} > 0. \quad (8.2)$$

Let Σ_k be the following set of $2n$ points

$$\Sigma_k = \{-(p_k^N)^i / (p_k^C)^i, -(q_k^N)^i / (q_k^C)^i, i = 1, 2, \dots, n\}$$

and define the distance from σ to the set Σ_k as

$$\text{dist}(\sigma, \Sigma_k) = \min\{|\sigma - \varsigma| : \varsigma \in \Sigma_k\}.$$

We choose σ_k according to Procedure 2 with the additional restriction that

$$\text{dist}(\sigma_k, \Sigma_k) \geq \pi_k \omega_k / (8n + 4). \quad (8.3)$$

In other words, we require not only

$$\sigma_k \in [0.5(\rho^l + \rho_k^u)\omega_k, \rho_k^u \omega_k] \quad (8.4)$$

but also that σ_k be bounded away from the set Σ_k by at least $\pi_k \omega_k / (8n + 4)$. Since $\{\pi_k\}$ is bounded away from zero, we see from (8.3) that $\{\text{dist}(\sigma_k, \Sigma_k)\}$ is bounded away from zero if $\{\omega_k\}$ is bounded away from zero.

The purpose of introducing condition (8.3) is to avoid the situation where $p_k^i = (p_k^N)^i + \sigma_k (p_k^C)^i$ (say) converges to zero but $(p_k^N)^i$ and $(p_k^C)^i$ do not. Although we believe that this situation is extremely unlikely to happen, we have not been able to rule it out.

Lemma 8.1 *The set of σ_k 's satisfying (8.3) and (8.4) is nonempty.*

Proof: The length of the interval in (8.4) is $\pi_k \omega_k / 2$. Partition this interval into $2n + 1$ equal sub-intervals, each having length $\pi_k \omega_k / (4n + 2)$. If the interior of any one of the

sub-intervals does not intersect Σ_k , then the midpoint of that sub-interval will satisfy (8.3) and (8.4). Since Σ_k has only $2n$ points, it cannot intersect the interiors of all the $2n + 1$ sub-intervals. This proves the lemma. \square

Now we are well-equipped to prove our superlinear convergence theorem.

Theorem 8.1 (Superlinear Convergence)

Let (x_, y_*) be a solution of problem (1.3) and $\{(x_k, y_k)\}$ be generated by Algorithm 2 with the restriction (8.3) on the centering parameter σ_k . Assume*

- (i) *strict complementarity holds at (x_*, y_*) ,*
- (ii) *ρ^u is sufficiently large, e.g., $\rho^u \geq 16\Gamma$.*

If $\{(x_k, y_k)\}$ converges to (x_, y_*) , then the duality gap sequence $\{x_k^T y_k\}$ converges to zero Q -superlinearly.*

Proof: We first prove $\sigma_k \rightarrow 0$. It suffices to show $\omega_k \rightarrow 0$.

Let $x_*^i > 0$. Obviously,

$$1 = \lim_{k \rightarrow \infty} \frac{x_{k+1}^i}{x_k^i} = \lim_{k \rightarrow \infty} (1 + \alpha_k p_k^i).$$

This implies $p_k^i \rightarrow 0$, because $\{\alpha_k\}$ is bounded away from zero. On the other hand, if $x_*^i = 0$, then $y_*^i > 0$ by strict complementarity. The same argument, interchanging the roles of p_k^i and q_k^i , gives $q_k^i \rightarrow 0$. Therefore, for each index i , either

$$p_k^i = (p_k^N)^i + \sigma_k (p_k^C)^i \rightarrow 0 \quad \text{or} \quad q_k^i = (q_k^N)^i + \sigma_k (q_k^C)^i \rightarrow 0. \quad (8.5)$$

We will prove $\omega_k \rightarrow 0$ by contradiction. Suppose the opposite. Then, there exists a subsequence $\{\omega_{k_0}\} \subset \{\omega_k\}$ that is bounded away from zero. This in turn implies, from (8.3), that $\{\text{dist}(\sigma_{k_0}, \Sigma_{k_0})\}$ is bounded away from zero (recall that π_k is bounded away from zero).

We have shown that for each index i , either $p_k^i \rightarrow 0$ or $q_k^i \rightarrow 0$. Without loss of generality, assume $p_k^i \rightarrow 0$. We now show that $\{(p_{k_0}^C)^i\}$ converges to zero. Otherwise,

there exists a subsequence $\{(p_{k_1}^C)^i\} \subset \{(p_{k_0}^C)^i\}$ such that $\{|(p_{k_1}^C)^i|\}$ is bounded away from zero. For this subsequence,

$$p_{k_1}^i = (p_{k_1}^N)^i + \sigma_{k_1}(p_{k_1}^C)^i = (p_{k_1}^C)^i[\sigma_{k_1} + (p_{k_1}^N)^i/(p_{k_1}^C)^i] \rightarrow 0.$$

This implies

$$\sigma_{k_1} + (p_{k_1}^N)^i/(p_{k_1}^C)^i \rightarrow 0.$$

However, this cannot be true because $\{\text{dist}(\sigma_{k_1}, \Sigma_{k_1})\} \subset \{\text{dist}(\sigma_{k_0}, \Sigma_{k_0})\}$ is bounded away from zero. Hence, $(p_{k_0}^C)^i \rightarrow 0$.

Now in view of (8.5) we also have $(p_{k_0}^N)^i \rightarrow 0$. Similarly, we can prove that if $q_k^j \rightarrow 0$, then we have both $(q_{k_0}^N)^j \rightarrow 0$ and $(q_{k_0}^C)^j \rightarrow 0$. Therefore, for each index i , either $(p_{k_0}^N)^i$ and $(p_{k_0}^C)^i$, or $(q_{k_0}^N)^i$ and $(q_{k_0}^C)^i$ converge to zero. Since all these sequences are uniformly bounded (see the proof of Lemma 4.1), this leads to $\omega_{k_0} \rightarrow 0$ (see definition (4.2)), contradicting the hypothesis that $\{\omega_{k_0}\}$ is bounded away from zero. This proves that $\omega_k \rightarrow 0$. Consequently $\sigma_k \rightarrow 0$.

Now we prove $\alpha_k \rightarrow 1$. Note that (2.4) can be written as

$$p_k + q_k = -e + \sigma_k \frac{1}{n} x_k^T y_k (X_k Y_k)^{-1} e.$$

Since $\frac{1}{n} x_k^T y_k (X_k Y_k)^{-1} e$ is bounded above by $1/\gamma$, as $\sigma_k \rightarrow 0$, we have

$$p_k + q_k \rightarrow -e.$$

We have shown that for each i , either $p_k^i \rightarrow 0$ or $q_k^i \rightarrow 0$. Therefore, all p_k^i and q_k^i converge to either 0 or -1 . This again implies that $\hat{\alpha}_k \rightarrow 1$ (see (7.3)). In view of (6.1) and (7.2), α_k will eventually be equal to or greater than one if ρ^u is sufficiently large, e.g., $\rho^u \geq 16\Gamma$. Hence,

$$1 \leq \alpha_k \leq \hat{\alpha}_k \rightarrow 1.$$

This completes the proof. □

9 Concluding Remarks

In this paper, we have shown that the two fundamental parameters in primal-dual interior-point algorithms for linear programming can be chosen in such a way that both polynomiality and superlinear convergence are achieved. If the solution is a nondegenerate vertex, then in addition to superlinear convergence we have quadratic convergence.

The current practices in some of the state-of-the-art implementations of primal-dual interior-point algorithms have the following common fundamental features. First, they allow iterates to be very close to the boundary of the positive orthant; second, they phase out the centering steps at a fast pace. The theory established in Zhang, Tapia and Dennis [11] has already provided theoretical justification for such a practice from the viewpoint of fast convergence. This paper provides further theoretical justification for such a practice from the viewpoint of polynomiality. In summary, one can indeed, under reasonable conditions, accomplish both objectives — good global behavior and good local behavior.

We recently learned of a new result by Güler and Ye [4]. When applied to linear programming, it says that condition (3.1) will guarantee strict complementarity for any limit point of the iteration sequence generated by an interior-point algorithm. This result nicely complements the Zhang-Tapia-Dennis theory (i.e., Theorems 2.1 and 2.2) and therefore, the strict complementarity assumptions in Theorems 7.1 and 8.1 are no longer necessary.

Acknowledgment

We gratefully acknowledge Florian Potra and Michael Todd for their useful comments on an earlier version of this paper. We also thank Yinyu Ye and two anonymous referees for their constructive suggestions.

References

- [1] I. C. Choi, C.L. Monma, and D.F. Shanno. Further development of a primal-dual interior point method for linear programming. *ORSA J. on Computing* 2:304-311, 1990.
- [2] M. Iri and H. Imai. A multiplicative barrier function method for linear programming. *Algorithmica* 1:455-482, 1986.
- [3] M. Kojima, S. Mizuno, and A. Yoshise. A primal-dual interior point method for linear programming. In Nimrod Megiddo, editor, *Progress in Mathematical Programming, interior-point and related methods*, pages 29–47. Springer-Verlag, New York, 1989.
- [4] O. Güler and Yinyu Ye. Convergence behavior of some interior-point algorithms. Working paper series No. 91-4, the College of Business Administration, The University of Iowa, 1991.
- [5] I.J. Lustig, R.E. Marsten, and D.F. Shanno. Computational experience with a primal-dual interior point method for linear programming. *Journal of Linear Algebra and Applications*, 152:191–222, 1991.
- [6] K.A. McShane, C.L. Monma, and D.F. Shanno. An implementation of a primal-dual interior point method for linear programming. *ORSA J. on Computing*, 1:70–83, 1989.
- [7] R.C. Monteiro and I. Adler. Interior path-following primal-dual algorithms. Part I: linear programming. *Math. Prog.*, 44:27–41, 1989.
- [8] S. Mizuno, M.J. Todd and Y. Ye. On adaptive-step primal-dual interior-point algorithms for linear programming. Technical Report No. 944, School of Operations

- Research and Industrial Engineering, Cornell University, 1990. To appear in *Math. of O.R.*.
- [9] M.J. Todd and Y. Ye. A centered projective algorithm for linear programming. *Math. of O.R.*, 15:508–529, 1990.
 - [10] H. Yamashita. A polynomially and quadratically convergent method for linear programming. Report, Mathematical Systems Institute, Inc. 10F Ohashi-Daini Bldg., 2-4-3 Shinjuku, Shinjuku-ku, Tokyo, Japan, 1986.
 - [11] Y. Zhang, R. A. Tapia, and J. E. Dennis. On the superlinear and quadratic convergence of primal-dual interior-point linear programming algorithms. Technical Report TR90-6, Dept. Mathematical Sciences, Rice University, 1990. To appear in *SIAM J. Optimization*.